



UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PESQUISA

PROGRAMA DE INICIAÇÃO CIENTÍFICA VOLUNTÁRIA – PICVOL

**AVALIAÇÃO DA COMPLEXIDADE ESPAÇO-TEMPORAL DA
PRECIPITAÇÃO PLUVIAL NO BRASIL**

**AVALIAÇÃO COMPLEXIDADE ESPAÇO-TEMPORAL DO NÚMERO DE
DIAS COM PRECIPITAÇÃO NO BRASIL**

Área do Conhecimento: Ciências Exatas

Subárea do Conhecimento: Estatística

Especialidade do Conhecimento: Probabilidade e Estatística Aplicada

Relatório Final

Período: de Agosto de 2019 a Julho de 2020

PICVOL

SUMÁRIO

| | |
|--|----|
| 1. INTRODUÇÃO | 3 |
| 2. OBJETIVOS | 4 |
| 2.1. Objetivo Geral | 4 |
| 2.2. Objetivos específicos | 4 |
| 3. METODOLOGIA..... | 5 |
| 3.1. Levantamento dos Dados | 5 |
| 3.2. Sample Entropy (SampEn) | 5 |
| 3.3. Ponderação pelo Inverso da Distância (IDW) | 7 |
| 4. RESULTADOS E DISCUSSÕES | 8 |
| 5. CONCLUSÕES | 14 |
| 6. PERSPECTIVAS DE FUTUROS TRABALHOS | 15 |
| 7. REFERÊNCIAS BIBLIOGRÁFICAS | 16 |
| 8. OUTRAS ATIVIDADES | 18 |
| 9. JUSTIFICATIVA DE ALTERAÇÃO NO PLANO DE TRABALHO | 19 |

1. INTRODUÇÃO

Uma das preocupações quanto às chuvas é a intensidade e a frequência de suas ocorrências, pelos seus efeitos potencialmente danosos, quando em excesso ou por escassez. O conhecimento das probabilidades de ocorrência de chuva é de suma importância no planejamento agrícola, possibilitando o plantio em época adequada, o planejamento de irrigação, aplicação de adubos e defensivos, a utilização de máquinas agrícolas, etc. É por meio da agrometeorologia operacional que se pode obter valores de estimativas que irão permitir tomadas de decisão, tais como: preparo do solo, semeadura, irrigação e colheita (Silva, 2014).

As informações utilizadas para a determinação do clima são obtidas principalmente de estações meteorológicas que registram as variáveis atmosféricas (temperatura do ar, umidade relativa do ar, pressão atmosférica, precipitação entre outras) (Reboita et. al, 2015). Estas informações que são coletadas, gerenciadas e analisadas ajudam tomadores de decisão e usuários a planejar e adaptar suas atividades e projetos às condições esperadas. Desta maneira, decisões podem ser tomadas no planejamento, o que reduz riscos e aperfeiçoam os benefícios sócio-econômicos. (INMET).

O estudo do clima é de suma importância para compreender a dinâmica climática, analisar as possibilidades de chuva, prevendo dessa forma prováveis deslizamentos e desastres naturais, permitindo que medidas sejam tomadas previamente.

No decorrer deste relatório será feita a análise dos dados referente ao número de dias com precipitação no Brasil, sendo que, ao final do estudo poderá ser indicado regiões com baixa e alta complexidade, para assim se ter uma noção em quais regiões serão mais fácil de realizar modelagens e previsões, quanto a quantidade de dias em que chovem.

2. OBJETIVOS

2.1. Objetivo Geral

Avaliar a complexidade espaço-temporal do número de dias com precipitação no Brasil utilizando a estatística *Sample Entropy*.

2.2. Objetivos específicos

- Realizar a análise exploratória de dados total e mensal.
- Inferir a complexidade espaço-temporal da precipitação acumulada no Brasil.
- Realizar a interpolação de espacial das estatísticas obtidas.
- Analisar se existe padrões de variabilidade espacial nos dados.
- Identificar possíveis fatores correlacionado.

3. METODOLOGIA

3.1. Levantamento dos Dados

Os dados utilizados se referem a séries históricas diárias de 264 estações meteorológicas, distribuídas no território brasileiro, compreendendo o período de 1998 à 2018, sendo essas informações obtidas pelo Instituto Nacional de Meteorologia (INMET). Estes dados são disponibilizados para download pelo INMET no site: <<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>>.

Para resolver o problema em que há meses no ano com quantidade de dias diferentes, utilizamos a variável proporção de dias com chuva no mês. Essa proporção varia entre 0 e 1, onde o 0 representa que em não houve registro de chuva em nenhum dia do mês, e o 1 representa que houve registro de chuva em todos os dias do mês. Os meses em que houveram mais de 15% de seus dias sem registros se houve precipitação, foram retirados das séries temporais.

As análises descritivas dos dados e a estatística Sample Entropy foram feitas no software R (Project for Statistical Computing) versão 3.6.1. A interpolação espacial foi feita com o método IDW e a plotagem dos mapas foram realizadas com o software Surfer versão 11.0.

3.2. Sample Entropy (SampEn)

O método *SampEn* introduzido por Richman e Moorman (Richman e Moorman, 2000) é definido como sendo o logaritmo natural da probabilidade condicional de que duas sequências similares (dentro de um nível de tolerância r) para m pontos, permanecem similares para $m + 1$ pontos.

Seja $x(1), \dots, x(N)$ uma série temporal de tamanho N , o algoritmo de *SampEn* (m, r, N) é descrito como:

- i) Construir $N - m$ vetores de tamanho m , em que

$$x(i) = [x(i), x(i + 1), \dots, x(i + m - 1)], i = 1, \dots, N - m + 1; \quad (3.2.1)$$

- ii) Definir a distância d_m entre os vetores $x(i)$ e $x(j)$, onde

$$d_m[x(i), x(j)] = \max_{k=0, m-1} [|x(i + k) - x(j + k)|]; \quad (3.2.2)$$

- iii) Para cada $i = 1, \dots, N - m$; calcula-se

$$B_i^m(r) = \frac{B_i}{(N-m-1)}, \quad A_i^m(r) = \frac{A_i}{(N-m)}; \quad (3.2.3)$$

onde B_i é o número dos vetores $x(j)$ de tamanho m que são similares aos vetores $x(i)$ dentro da distância r : $d_m[x(i), x(j)] \leq r$ e A_i é o número dos vetores $x(j)$ de tamanho $(m + 1)$ que são similares aos vetores $x(i)$ dentro da distância r .

iv) Calcula-se

$$\begin{aligned} B^m(r) &= \frac{1}{N - m} \left(\sum_{i=1}^{N-m} B_i^m(r) \right), \\ A^m(r) &= \frac{1}{N - m} \left(\sum_{i=1}^{N-m} A_i^m(r) \right), \end{aligned} \quad (3.2.4)$$

onde $B^m(r)$ é a probabilidade de dois vetores serem similares para m pontos, $A^m(r)$ é a probabilidade de dois vetores serem similares para $(m + 1)$ pontos.

v) Por fim, calcula-se o índice *Sample Entropy*:

$$SampEn(m, r, N) = -\ln \left(\frac{A^m(r)}{B^m(r)} \right), \quad (3.2.5)$$

que representa a probabilidade condicional que de duas sequências que são similares para m pontos permanecem similares para $m + 1$ pontos (quando um dado consecutivo é adicionado).

A escolha do parâmetro m define o tamanho dos vetores inicialmente formados no algoritmo. Quando escolhemos o valor $m = 2$, por exemplo, estamos escolhendo mensurar as flutuações existentes em cada par de observações. Através do parâmetro r é que definimos se essa diferença em cada par será considerada significativa ou não. O parâmetro r usualmente é utilizado como percentual do desvio padrão dos dados. Por exemplo, se nossa série possui desvio padrão igual a 10 u.m., e se adotarmos $r = 0,2$ e $m = 2$, estamos assumindo em nosso algoritmo que diferenças maiores que 2 u.m., entre cada par de observações, será considerada como significativa. Quão menor for o valor de r , maior será o rigor na estimativa da entropia. Em uma mesma série temporal, a variação de r fará com que tenhamos maiores estimativas de *SamEn* para os maiores valores de r adotados.

Sample Entropy é uma modificação do método *Approximate Entropy* (*ApEn*) (Pincus e Viscarello, 1992). Ambos os métodos servem para quantificar a complexidade em séries temporais não lineares, que foram amplamente usados em análise de processos fisiológicos (Tang et al., 2004) e séries climáticas (Shuangcheng et al., 2006). Uma maior complexidade (menor

regularidade) da série temporal gera o valor do *SampEn* mais alto (Richman e Moorman, 2000). De acordo com Shuangcheng et al. (Shuangcheng *et al.*, 2006), a interpretação dos valores da estatística de *SampEn* considera uma série totalmente regular quando *SampEn* = 0 e para uma série totalmente irregular, *SampEn* = 3.

3.3. Ponderação pelo Inverso da Distância (IDW)

O IDW é uma técnica de interpolação muito utilizada para quando se tem pontos espalhados espacialmente, utilizando uma combinação linear ponderada para determinar os pontos amostrados. O peso imputado a cada ponto é em função do inverso da distância entre as amostras (MARCUIZZO, ANDRADE e MELO, 2011).

Para calcularmos o IDW, tem-se a seguinte fórmula:

$$Z(x) = \frac{\sum_{i=1}^n w_i Z(x_i)}{\sum_{i=1}^n w_i} \quad (3.3.1)$$

onde:

- $Z(x)$ – é o valor do ponto que deseja-se interpolar
- n – é o número de pontos próximos usados para interpolar o ponto x .
- $Z(x_i)$ – é o valor resultante do ponto x_i
- w_i – é o peso do ponto x_i sobre o valor de x .

A fim de encontrar o valor de w_i , utiliza-se equação a seguir:

$$w_i = \frac{1}{h(x, x_i)^p} \quad (3.3.2)$$

sendo:

- $h(x, x_i)$ – a distância entre o ponto x e o ponto x_i
- p – é o parâmetro da potência, comumente utiliza-se este valor igual a 2.

Parâmetros de potência maiores enfatizam pontos mais próximos, tornando o resultado menos suave. Parâmetros de potência menores enfatizam pontos mais distantes, tornando o resultado mais suave, porém menos preciso (MARCUIZZO, ANDRADE e MELO, 2011).

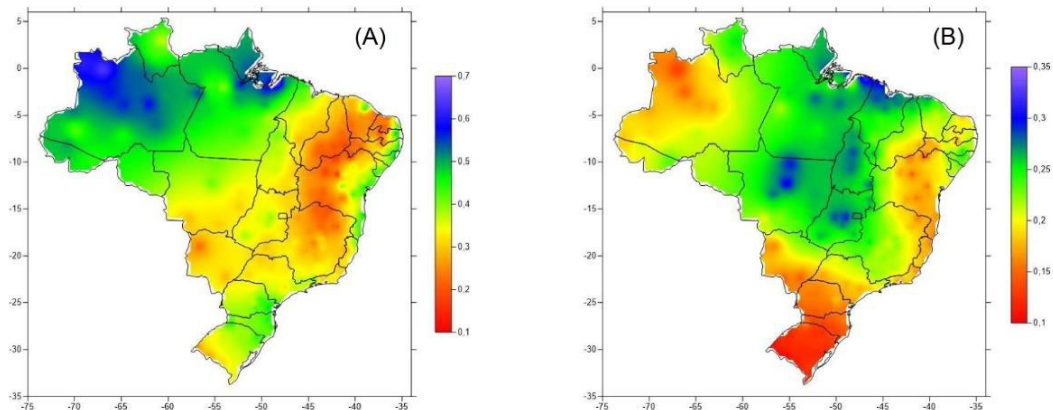
4. RESULTADOS E DISCUSSÕES

Os resultados encontrados nesse estudo para as médias e desvios padrão das observações, além da estatística Sample Entropy, foram interpolados e distribuídos espacialmente no mapa do Brasil, de acordo com a localização de cada uma das 264 estações meteorológicas, para melhor visualização e entendimento dos resultados.

A média do número de dias com precipitação (Figura 1 A) mostrou que os valores maiores se encontram na região Norte, mais especificamente na área que compreende a Amazônia e a zona climática equatorial que tem a característica a alta ocorrência de chuvas durante todo o ano, a explicação para isso, é o fato de que a região amazônica possui um fenômeno chamado de “rios voadores”, que devido a grande quantidade de árvores, a floresta amazônica serve como uma bomba d’água, puxando para dentro do continente a umidade evaporada pelo oceano Atlântico e carregada pelas massas de ar, com isso a umidade cai como chuva sobre a floresta e pela ação da evapotranspiração, a floresta devolve a água da chuva para a atmosfera na forma de vapor de água, dessa maneira, o ar é sempre recarregado com mais umidade. (RIOS VOADORES, 2013) Os menores valores ficaram na região Nordeste, Sudeste e Centro-Oeste, principalmente nos biomas da Caatinga e do Cerrado, onde previamente sabemos que a ocorrência de chuvas se dá em determinado período do ano e no restante é quase nula essa eventualidade.

Quanto ao desvio padrão (Figura 1 B) verificou a região Sul apresentou os menores desvios mostrando que esta possui uma ocorrência de chuvas distribuída da mesma forma ao longo do período observado, a região Nordeste também apresentou baixos desvios e anteriormente observou que esta região apresentou poucos dias com precipitação, então, isso significa que esta possui baixa quantidade de dias que chovem ao longo de todo o período estudado, já a região com maior variabilidade é compreendida por parte da região norte e do centro-oeste chegando a variar em 35%, que pode ser um indicativo de uma mudança na ocorrência de chuvas ao longo desses 20 anos.

Figura 1: Média (A) e desvio padrão (B) do número de dias com precipitação no Brasil, no período de 1998 a 2018.

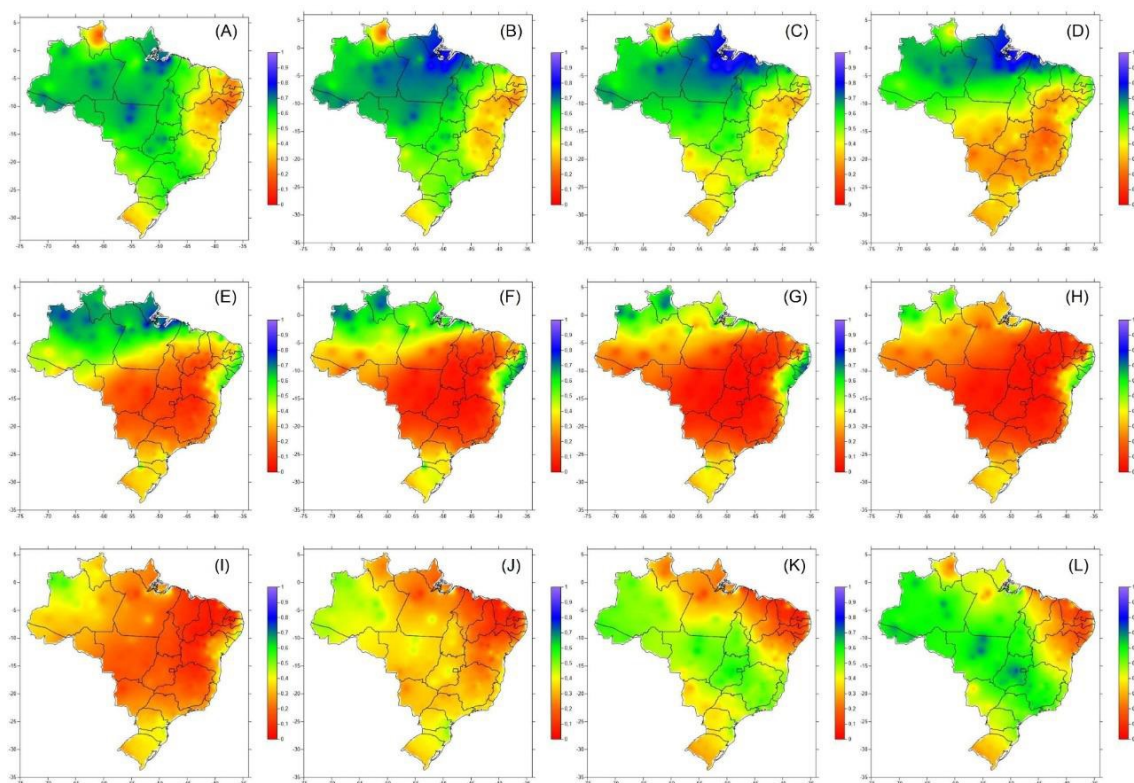


Fonte: Próprio Autor

Com o intuito de observar como se dá a evolução ao passar dos meses (Figura 2), no período de 1998 a 2018, do número de dias em que ocorreram precipitação, foram feitos gráficos com todos os 12 meses do ano, utilizando como métrica a média do número de dias de cada uma das 264 estações no determinado mês e para ter como comparar os meses foi criada uma escala que o intervalo desta vai de 0 a 1, onde o valor zero resulta que não teve nenhum dia com precipitação, em contra partida o número um representa que todos os dias tiveram chuva.

Nesta figura nota-se que nas regiões Norte, Centro-Oeste e Sudeste, chove em mais da metade dos dias nos dois primeiros meses do ano: janeiro (A) e fevereiro (B). Em março (C) verifica-se que na região Sudeste há uma diminuição na quantidade de dias em que ocorre precipitação, enquanto nas regiões Centro-Oeste e Norte, continuam tendo uma alta ocorrência de chuvas ao longo dos dias, contudo em abril (D) verificamos uma queda na quantidade de dias com chuvas, também na região Centro-Oeste, neste mês temos um cenário onde a região Norte é a única que continua chovendo em maior parte do mês, e temos também boa parte do litoral nordestino com esta característica em abril ao longo do período observado.

Figura 2: Média do número de dias com precipitação por mês, no período de 1998 a 2018.



Fonte: Próprio Autor

Nos quatro meses subsequentes, maio (E), junho (F), julho (G), agosto (H), verifica-se uma tendência ao longo de todo o território brasileiro em termos menos dias com precipitação em relação aos quatro meses iniciais, tendo maior parte do território apresentando pouquíssimos dias nesses meses com precipitação, porém o litoral nordestino no segundo quadrimestre nos mostra que chove em aproximadamente metade dos dias desses meses, consequentemente, temos que, enquanto o restante do território brasileiro tem uma tendência em diminuir a quantidade de dias com precipitação nos quatro meses em questão, o litoral nordestino apresenta tem a tendência em aumentar o número de dias em que ocorrem chuvas. Nos próximos dois meses do ano, setembro (I) e outubro (J), observa-se que são os meses com menos dias com precipitação, chegando a valores próximos ao zero da escala, em todo o Brasil. No mês de novembro (K), verifica-se que em maior parte das regiões Centro-Oeste e Sudeste há um aumento na quantidade de dias com chuvas, chegando a ter novamente aproximadamente metade do mês com precipitação, enquanto

isso o Nordeste apresenta baixíssimas quantidade de dias em que chovem. Finalmente, em dezembro (L), nos é mostrado que as regiões Norte, Centro-Oeste e Sudeste, apresentam maior parte de seus territórios com chuvas em aproximadamente metade deste mês, enquanto novamente neste o Nordeste apresenta as menores quantidades de dias em que ocorre precipitação em todo o Brasil.

A região Sul foi a região mais “constante” ao longo do ano, apresentando somente, em janeiro (A) e fevereiro (B), em maior parte de seu território, uma quantidade de dias em torno da metade do mês, em que houve precipitação, no restante do ano, observa-se que esta região, apresentou quantidades inferiores a metade do mês, em que choveram.

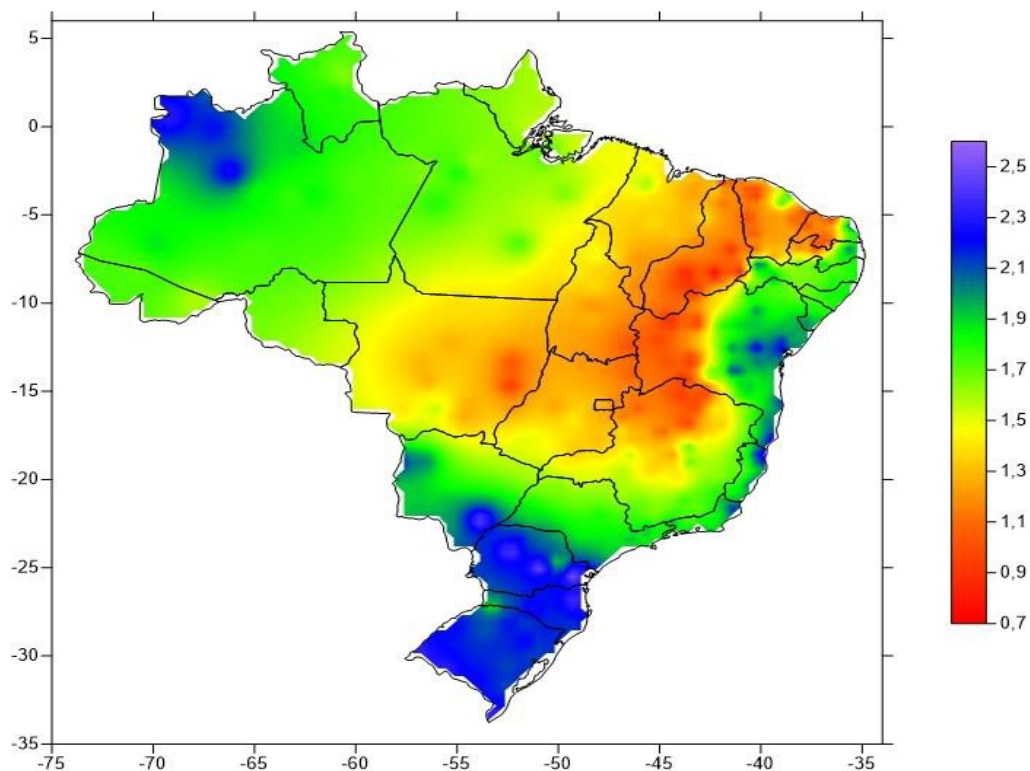
Para verificar a complexidade das séries temporais referentes ao número de dias com precipitação de cada uma das 264 estações, distribuídas em todo o Brasil, no período compreendido de 1998 a 2018, foi calculado a entropia dessas séries, utilizando as estimativas resultadas da técnica Sample Entropy. Onde está estatística mede a complexidade de nossas séries temporais, nos dando um grau da desordem destas. Esta medida varia no intervalo de 0 a 3, sendo que o valor 0 mostra que temos uma série regular e o 3 indica que temos uma série totalmente irregular. Todas as estimativas da medida Sample Entropy neste estudo, foram feitas utilizando $m = 2$, $r = 0,2$.

Os valores da estatística Sample Entropy foram distribuídos espacialmente em todo o território brasileiro de acordo com cada estação meteorológica, que pode ser visto na Figura 3, nestes dados a escala ficou entre 0,7 e 2,6, mostrando que temos valores estimados de entropia bastante diferenciados no Brasil, tendo regiões apresentando valores baixos e outras o contrário, observa-se também que os valores da escala dos dados, ficaram próximos aos valores extremos da Sample Entropy, sendo zero o menor valor possível da técnica, donde, este indica uma série tendo um padrão claro e três como o limite superior que mostra que a série não tem nenhum padrão previsível.

Nota-se que os maiores valores da entropia está concentrado em toda a região Sul, e em parte dos estados do Amazonas, da Bahia e do Mato Grosso do Sul, mostrando que, principalmente, a região Sul apresentou uma grande variabilidade na quantidade de dias em que se terá precipitação ao longo do

período estudado. Já maior parte da região Centro-Oeste, unindo-se a parte das regiões Norte e Nordeste, formam o território onde apresentam as menores entropias em todo o mapa, sendo que dentre os estados o que apresentou as menores entropias em todo o seu território foi o estado do Piauí.

Figura 3: Coeficiente de Entropia (Sample Entropy) das séries com o número de dias com precipitação no Brasil, no período de 1998 a 2018.



Fonte: Próprio Autor

Observamos que a região Sul foi a região com os maiores índices de entropia, onde esta região apresenta os biomas Pampas e Mata Atlântica, estes parecem ter influenciado a essa região manifestar esses altos índices, pois verifica-se que em todo o território do bioma de Mata Atlântica é apresentado valores iguais e superiores a 1,7, mostrando que esses biomas tem alta variabilidade durante os meses quanto ao número de dias em que ocorre precipitação. Quanto a região Sul, verificou-se que no número médio de dias que chovem, de acordo com o mês (Figura Y), que nesta região quase todos os meses permaneceram contendo menos da metade dos dias com precipitação, nos mostrando que possivelmente a variação mensal durante os anos é o que torna esta região tão imprevisível. Outro bioma que apresentou taxas maiores

que a metade da escala, foi a Amazônia que tem como característica a grande quantidade de dias com chuvas nos dois últimos meses do ano e nos primeiros quatro meses, sendo eles, janeiro, fevereiro, março e abril, contudo no restante do ano a uma queda na quantidade de dias que chovem, e provavelmente essa queda não seja tão previsível assim, com a relação ao número médio de dias em que há precipitação. Já os biomas Cerrado e Caatinga, são os que possuem mais regularidade, segundo a medida de entropia, em termos de modelagem, nesses biomas, consegue-se fazer previsões de forma mais fácil, em comparação com os demais.

5. CONCLUSÕES

Conclui-se, que no Brasil, no período de janeiro de 1998 a dezembro de 2018, tem-se uma região Norte com a maior média dos dias em que ocorrem precipitação, em contra partida a região Nordeste detém as piores médias, exceto, no litoral nordestino, que fica com valores aproximados no meio da escala. Em relação a evolução mensal da quantidade de dias com precipitação, verificou-se que nos três primeiros meses e nos últimos dois meses do ano tem-se um número de dias em que se tem ocorrência de chuvas maior, enquanto que após o terceiro mês verifica-se uma tendência em maior parte do país, em diminuir a ocorrência de chuvas. Quanto a estatística Sample Entropy observou que no Sul do Brasil foram vistos os valores mais altos da Entropia em toda a região, mostrando que é difícil modelar o número de dias com precipitação nessa região, e os menores valores é observado no centro do país.

Com tudo, verificamos que a área dos biomas do Brasil delimita de melhor forma, do que as regiões do Brasil, onde o comportamento apresentado pelo número de dias em que há precipitação em alguns biomas é maior que a média durante o ano e em outros apresenta taxas baixíssimas.

Por fim, com estas análises podem ser feitos modelos nas regiões onde a entropia apresentou valores pequenos, ou seja, há padrões definidos, o que facilita a modelagem, para entender melhor o padrão climático destas regiões.

6. PERSPECTIVAS DE FUTUROS TRABALHOS

Com o conhecimento acerca da entropia do número de dias com precipitação em um mês, podemos utilizar estes resultados para melhorar modelos de previsão climática já existentes.

7. REFERÊNCIAS BIBLIOGRÁFICAS

SILVA, José Rodrigo Santos. **Avaliação de autocorrelações e complexidade de séries temporais climáticas no Brasil**. Recife - PE: Tese (Doutorado em Biometria e Estatística Aplicada) - Universidade Federal Rural de Pernambuco, 2014. 114 p.

REBOITA, M. S.; KRUSCHE, N.; AMBRIZZI, T.; ROCHA, R. P. DA. Entendendo o tempo e o clima na América do Sul. **Terrae Didactica**, v. 8, n. 1, p. 34, 29 jun. 2015.

INMET. Clima. Disponível em: <<http://www.inmet.gov.br/portal/index.php?r=home/page&page=clima>>. Acesso em: 15 de jul. de 2020.

RIOS VOADORES. **Fenômeno dos rios voadores**. Disponível em: <<https://riosvoadores.com.br/o-projeto/fenomeno-dos-rios-voadores/>>. Acesso em: 10 de jul. de 2020.

MARCUZZO, F. F. N.; ANDRADE, L. R. D.; MELO, D. C. R. Métodos de Interpolação Matemática no Mapeamento de Chuvas do Estado do Mato Grosso. **Revista Brasileira de Geografia Física**, v. 4, n. 4, p. 793-04, 2011. ISSN 194-2295.

RICHMAN, J S; MOORMAN, J R. Physiological time-series analysis using approximate entropy and sample entropy. **American journal of physiology. Heart and circulatory physiology** v. 278, n. 6, p. H2039-49 , jun. 2000.

PINCUS, S M; VISCARELLO, R R. Approximate entropy: a regularity measure for fetal heart rate analysis. **Obstetrics and gynecology** v. 79, n. 2, p. 249–55 , fev. 1992. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/1731294>>. Acesso em: 20 de fev. de 2020.

INMET. Dados históricos. Disponível em: <<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>>. Acesso em: 27 de out. de 2019.

TANG, Xiao-jun *et al.* Complexity measurements of electroencephalograph recordings using sample entropy algorithm in patients with temporal lobe epilepsy. **Acta Biophysica Sinica** v. 20, p. 382–392 , 2004.

SHUANGCHENG, Li *et al.* Measurement of climate complexity using sample

entropy. **International Journal of Climatology** v. 26, n. 15, p. 2131–2139 , dez.
2006.

8. OUTRAS ATIVIDADES

- Estágio em estatística na SMTT de Aracaju.
- Membro da ATUE, Empresa Júnior de Estatística da UFS

9. JUSTIFICATIVA DE ALTERAÇÃO NO PLANO DE TRABALHO

Não houve alterações no plano de trabalho.

